

**Amendments to the Specification:**

Please replace the paragraph beginning at page 5, line 29 with the following amended paragraph:

First, a phrase for which a definition is sought is provided (block [[310]] 41). The phrase may be provided by, for example, a user request or query, or by any other means. One example of a system for providing a phrase is [[that]] located at the URL identified by <http://labs.google.com/glossary>, the contents of which are incorporated by reference. In addition, the spelling of the phrase can be corrected if necessary or normalized into a common root form to provide more consistent definition results.

Please replace the paragraph beginning at page 6, line 6 with the following amended paragraph:

Documents that contain definitions are determined (block [[320]] 42). These documents may be determined in any number of ways. For example, such documents may be determined during Web-crawling or spidering performed by search engines in either real time or batch processing modes. Once a document is determined to contain definitions, the document (or information about the document, such as the document's URL) may be stored or remembered for future use. "Authoritative" sources for definitions may also be used, for example, documents associated with Web sites, such as <http://www.dictionary.com> the Web site [dictionary.com](http://www.dictionary.com).

Please replace the paragraph beginning at page 7, line 3 with the following amended paragraph:

The phrase for which definition is sought is then matched against the determined documents to return definitions (block [[330]] 43). The documents determined in this step (block [[330]] 43) may be parsed to identify occurrences of the phrase being sought and the phrase's associated definition. For example, definition containing documents may be organized with 'headwords,' or words that can be looked up in a dictionary form. There are various methods for identifying headwords and/or identifying definitions. In one embodiment of the invention, one or more of the following methods are used to parse apart documents, identify headwords, and/or return definitions:

- If the page uses `<dl>`, `<dt>` and `<dd>`, which are HTML tags used for specifying lists of definitions, the HTML mark up is relied upon to identify definitions, that is:

An example definition list

```
<dl>
  <dt>Headword 1
    <dd>This is the definition of Headword 1
  <dt>Headword 2
    <dd>This is the definition of Headword 2
  <dt>Headword 3
    <dd>This is the definition of Headword 3
</dl>
```

- HTML tags, such as `<p>`, `<tr>`, `<li>`, and `<br>`, may be treated as separators between successive definitions.
- White space or punctuation (.,:-) is eliminated at the beginning of definitions.
- Headwords may be identified by the fact that the headwords are surrounded by the HTML tags `<b>`, `<strong>`, `<em>`, `<code>`, or `<span>`.
- Lines that do not start with headwords are deleted.

- If there are fewer than N, for instance, N=5, definitions found in the document or page, all definitions in the document or page are discarded.

Please replace the paragraph beginning at page 8, line 1 with the following amended paragraph:

The parser does not need to be perfect at identifying all headwords and definitions. In one embodiment, due to the large number of definition-containing documents determined in the definition document determination step (block [[320]] 42), the parser is biased towards precision rather than thoroughness. In other words, the parser errs towards throwing entries away rather than keeping entries that may be incorrect because there are more than enough definitions to supply a satisfactory outcome. Similarly, in a further embodiment, the parser de-duplicates entries that are duplicative or merely cumulative of other entries.

Please replace the paragraph beginning at page 8, line 9 with the following amended paragraph:

One or more of the returned definitions are then provided (block [[340]] 44). In one embodiment, the returned definitions are ranked according to PageRank<sup>TM</sup> of the documents from which they are retrieved, according to the methods disclosed in U.S. Patent No. 6,285,999, cited above. The retrieved definitions may also be processed for presentation, such as by carrying out one or more of the following steps:

- Removing:
  - all HTML markup;
  - leading and trailing white space in both headword and definition;
  - all punctuation (.:;!?-) in the headword;

- all leading non-alpha and non-parenthesis in the headword and definition;
  - all trailing non-alphanumeric and non-parenthesis in the headword.
- Throw the definition away if:
  - the definition starts with "see;
  - the definition is a duplicate of one already retrieved.
- Capitalize the first letter the definition.